MINISTRY OF EDUCATION, CULTURE AND RESEARCH OF REPUBLIC OF MOLDOVA

TECHNICAL UNIVERSITY OF MOLDOVA

FACULTY OF COMPUTERS, INFORMATICS AND MICROELECTRONICS

DEPARTMENT OF SOFTWARE ENGINEERING AND AUTOMATICS

PW Lab2. HTTP, Caching and content negotiation.

Project report

Mentor: prof., Alexei Sersun

Students: Covtun Serghei, FAF-203

Chișinău, 2023

Abstract

Imagine, you're a developer at TwoLines GmbH, a startup from Zurich, Switzerland. The idea of the startup is pretty simple - you develop best-in-class browser experience, with privacy and security in mind. One day, you receive a message from the CTO, Mike.

```
Hey there!

Tomorrow we'll have a presentation for a really biiiig investor.
They say they don't need PPTs. They want at least some proof-of-concept for secure browsers for their toasters.
Can you share with me the executable we used last time for demo?

Cheers,
Mike
```

Yeah, the catch is toasters don't have any browser exeprience!
Nonetheless, you can't remember where you've saved the executable, so you start writing it from scratch.

Content

Introduction

Your task for this lab is:

1. You have to write a command line program, using [go2web](go2web) executable as a starting point;
2. The program should implement at least the following CLI:
```

 go2web -u <URL>         # make an HTTP request to the specified URL and print the response
 go2web -s <search-term> # make an HTTP request to search the term using your favorite search engine
and print top 10 results
 go2web -h            # show this help
```

3. The responses from request should be human-readable (e.g. no HTML tags in the output)

 4. The app has to look attractive.

 Link to Git repo: https://github.com/siorkis/PW/tree/main/lab2

Work

To implement this laboratory work, I made sure that in python there are right tools as CLI parser, HTML/JSON parser and support for TCP sockets. In the laboratory work nr. 2 have been done executable with '-s', links from search engine can be accessed as well.

```
$ python go2web.py -h

go2web.py -s <search-term>    make an HTTP request to search via google
go2web.py -h                  show this help
```

Fig 1. Help mechanism.

By typing `-h` command user will see list of all commands. Command `-s` + `key` will launch google search with provided key. Program will return top 10 results with ability to click on that link. Connection are established with `socket` using web scrapping technique.

```python
client = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
client.settimeout(1)
client.connect(("www.google.com", 80))
request = "GET /search?q=%s HTTP/1.1\r\nHost: www.google.com\r\n\r\n" % key
```

Fig 2.1 GET requests via sockets.

```
$ python go2web.py -s world
GET /search?q=world HTTP/1.1
Host: www.google.com


World - Wikipedia
https://en.wikipedia.org/wiki/World


World Health Organization (WHO)
https://www.who.int/


World Bank Group - International Development, Poverty ...
https://www.worldbank.org/en/home


The World Economic Forum
https://www.weforum.org/


World Trade Organization - Home page - Global trade
https://www.wto.org/


FIFA World Cup Qatar 2022™
https://www.fifa.com/fifaplus/en/tournaments/mens/worldcup/qatar2022


World of Tanks — Легендарная онлайн-игра про танки
https://worldoftanks.eu/ru/


World of Tanks - Legendary Online Multiplayer Tank Game
https://worldoftanks.eu/
```

Fig 2.2 Searching in google with key=world.

World Bank
https://www.worldbank.org › ... ⋮

World Bank Group - International Development, Poverty ...

With 189 member countries, the **World** Bank Group is a unique global partnership fighting poverty worldwide through sustainable solutions.

World Economic Forum
https://www.weforum.org ⋮

The World Economic Forum

4 days ago — The **World** Economic Forum is an independent international organization committed to improving the state of the **world** by engaging business, ...

World Trade Organization
https://www.wto.org ⋮

World Trade Organization - Home page - Global trade

The **World** Trade Organization (WTO) deals with the global rules of trade between nations. Its main function is to ensure that trade flows as smoothly, ...

World Food Programme
https://www.wfp.org ⋮

UN World Food Programme (WFP)

We are the **world's** largest humanitarian organization, saving lives in emergencies and using food assistance to build a pathway to peace, stability and ...

FIFA
https://www.fifa.com › fifaplus › tournaments › mens ⋮

FIFA World Cup Qatar 2022™

The FIFA **World** Cup Qatar 2022™ will be played from 20 November to 18 December.
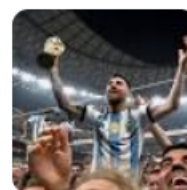
Fig 3. Actual search with word `world` in google.

Conclusion

Due to this laboratory work, I practiced building and working with HTTP requests, as well with parsing HTML responses. Thus I learned about new libraries that helped a lot in parsing process and considerably eased the creation of structured final output.

Bibliography

[1]   Internalpointers *Google*, https://www.internalpointers.com/post/making-http-requests-sockets-python

[2]   Datacamp *Google*, https://www.datacamp.com/tutorial/making-http-requests-in-python

[3]   Edureka *Google*, https://www.edureka.co/blog/web-scraping-with-python/